Synchronisation de Séquences d'Images Indépendamment des Points de Vue

Emilie Dexter, Patrick Pérez, Ivan Laptev

Synchronisation de Séquences d'Images Indépendamment des Points de Vue

View-independent Synchronization of Image Sequences

E. Dexter P. Pérez I. Laptev

IRISA / INRIA Centre Rennes Bretagne Atlantique

Campus Universitaire de Beaulieu, 35042 Rennes Cedex, France
emilie.dexter@irisa.fr

Résumé

Cet article aborde deux aspects de la synchronisation de séquences d'images : la synchronisation d'actions humaines et la synchronisation de scènes dynamiques. Le premier consiste à synchroniser différentes réalisations d'une même classe d'action alors que pour le second il s'agit du même évènement dynamique. Pour réaliser ces deux tâches indépendamment des points de vue, nous utilisons les auto-similarités temporelles comme descripteur de séquence. Bien que de tels attributs ne soient pas strictement invariants aux changements de vue, ils restent suffisamment stables. La synchronisation consiste alors en l'alignement de leurs descripteurs temporels par programmation dynamique. La méthode proposée a été validée sur plusieurs séquences d'images avec de grandes variations de points de vue, des mouvements de caméras et une grande variabilité intra-classe pour les actions humaines.

Mots Clef

Synchronisation, Matrice d'auto-similarité

Abstract

In this paper, we consider two instances of the problem of the image sequence synchronization : synchronization of human actions and synchronization of dynamic scenes. The first task consists of synchronizing different performances of action classe in video. The second one is to synchronize sequences representing the same dynamic event from different views. To address both tasks in a large view variation context, we use temporal self-similarities of sequences as temporal descriptors. Although such descriptors are not strictly invariant under view changes, they remain stable. Synchronizing two sequences is then performed by aligning their temporal descriptors by dynamic programming. The proposed method is validated on several videos with large view variations, drastic independent camera motions and the within-class variability of human actions.

Keywords

Synchronization, Self-similarity Matrix

1 Introduction

Synchroniser des séquences d'images se révèle une tâche nécessaire et critique lorsque l'alignement temporel n'est pas connu, pour des applications telles que la synthèse de nouvelles vues, la reconstruction $3D$ de scènes dynamiques ou la comparaison de séquences avec des contenus dynamiques similaires mais avec des variations de vitesse d'exécution. L'une des difficultés majeures réside dans l'ampleur des changements d'apparence. Ces changements peuvent être le fait de points de vue différents, de mouvements de caméra ou encore d'apparences changeantes des objets en mouvement.

Dans ces travaux, nous nous intéressons à deux problématiques de synchronisation dans un cadre commun d'indépendance aux changements de vue :

(i) Synchronisation d'actions, c'est-à-dire synchroniser des séquences représentant des réalisations différentes d'une même action

(ii) Synchronisation de vidéos, c'est-à-dire synchroniser des séquences d'un même évènement dynamique capté sous des points de vue différents.

Considérant deux séquences d'un même évènement dynamique, correspondant à des scènes identiques ou similaires, mais différant au moins par le point de vue, la synchronisation consiste en la mise en correspondance d'images de la première séquence avec des images de la seconde.

1.1 Travaux antérieurs

La synchronisation de vidéos a généralement été étudiée sous des hypothèses de caméras statiques et de transformation linéaire de l'axe temporel. Certains travaux estiment conjointement les transformations spatiale et temporelle entre deux séquences d'images [9, 3, 13] alors que d'autres ne s'intéressent qu'à l'aspect temporel [8, 2, 16, 14]. La majorité des approches exploitent des correspondances spatiales entre les vues, soit pour estimer une matrice fondamentale [3], soit pour utiliser des contraintes de rang sur des matrices d'observations [16]. D'autres méthodes essaient d'extraire des caractéristiques temporelles sans cor-

respondance, comme dans [14] où les auteurs explorent des descripteurs temporels basés sur les co-occurrences des changements d'apparence.

La synchronisation de séquences issues de caméras mobiles est peu abordée dans la littérature, et plus rarement encore de manière automatique. Par exemple dans [12], les auteurs choisissent manuellement 5 points mobiles indépendants afin de s'assurer que ces points soient suivis avec succès tout au long des séquences.

Toutes les approches mentionnées ci-dessus traitent du problème de synchronisation de séquences d'images de la même scène dynamique. Seuls deux de ces méthodes ont également été proposées pour aligner temporellement une même action réalisée par des personnes différentes [8, 14]. La première évalue l'alignement temporel en utilisant un algorithme de programmation dynamique connu sous le nom de Dynamic Time Warping (DTW) ainsi que des contraintes de rang sur des matrices d'observations exploitant des correspondances spatiales entre les séquences d'images. De telles correspondances sont en pratique difficiles à obtenir. La seconde approche estime les transformations spatiale et temporelle par la maximisation de corrélations spatiaux-temporelles locales pour des séquences d'images enregistrées par des caméras statiques ou se déplaçant conjointement.

1.2 Approche proposée

Dans cet article, nous proposons une approche originale de synchronisation automatique soit d'actions humaines soit de vidéos de la même scène dynamique enregistrées avec des points de vue substantiellement différents. Notre méthode combine une description temporelle des séquences d'images comme dans [14], ce qui ne requiert pas de correspondances de points entre les vues, avec de la programmation dynamique. L'algorithme DTW nous permet de traiter des déformations temporelles non-linéaires sous des contraintes de monotonie, ce qui est particulièrement intéressant lors de la synchronisation d'actions. Nous supposons seulement, pour l'instant, que les deux séquences montrent deux points de vue d'un même évènement dynamique ou d'une même classe d'action humaine.

Contrairement à la plupart des méthodes existantes, nous n'imposons pas d'hypothèse restrictive telle que des correspondances de points ou une connaissance a priori du type de "déformation" temporelle. Nous utilisons les matrices d'auto-similarité (SSM) comme descripteurs temporels des séquences d'images. Ce descripteur, présenté récemment dans [4] pour de la reconnaissance d'actions, n'est pas strictement invariant aux changements de vues mais relativement stable. En conséquence, des événements dynamiques similaires produisent des SSMs semblables. Les descripteurs temporels issus de ces matrices peuvent être mis en correspondance suivant une fonction de déformation temporelle que nous estimons par DTW. Le résultat se présente donc sous la forme d'une liste exhaustive de correspondances temporelles entre les séquences d'images.

La suite de l'article s'organise comme suit. La Section 2 introduit les descripteurs pour les séquences d'images ainsi que leur alignement temporel. Les Sections 3 et 4 sont consacrées aux résultats de synchronisation, respectivement pour les scènes dynamiques et pour les actions humaines. Enfin, nous concluons et proposons des directions pour les travaux futurs.

2 Méthode de synchronisation

Cette Section est consacrée à la présentation de la méthode commune permettant de synchroniser aussi bien des actions humaines que des scènes dynamiques. En premier lieu, nous présentons les descripteurs temporels pour les séquences d'images basés sur les auto-similarités temporelles. Ensuite, nous décrivons l'algorithme de Dynamic Time Warping (DTW) utilisé afin d'aligner ces descripteurs.

2.1 Descripteurs temporels

Le calcul des descripteurs temporels requiert deux étapes : (i) la construction pour chacune des séquences d'une matrice d'auto-similarité (SSM) qui caractérise les similarités et dissimilarités le long d'une séquence et (ii) le calcul d'un descripteur local qui capte les structures principales de cette matrice.

Matrices d'auto-similarités. Considérant une séquence d'images $I = \{I_1, I_2, \dots I_T\}$, la matrice d'auto-similarité, notée $\mathcal{D}(I)$, est exprimée par :

$$\mathcal{D}(I) = [d_{ij}]_{i,j=1\dots T} = \begin{bmatrix} 0 & d_{12} & \dots & d_{1T} \\ d_{21} & 0 & \dots & d_{2T} \\ \vdots & \vdots & & \vdots \\ d_{T1} & d_{T2} & \dots & 0 \end{bmatrix} \quad (1)$$

Il s'agit d'une matrice carrée symétrique où chaque élément d_{ij} représente une distance entre des attributs extraits des images I_i and I_j respectivement. Comparer les attributs d'une image avec elle-même conduit à une valeur zéro pour chaque élément diagonal associé.

La structure de cette matrice $\mathcal{D}(I)$ dépend des attributs choisis mais aussi de la distance considérée. Dans ces travaux, nous utilisons la distance euclidienne avec deux types d'attributs dynamiques afin de calculer les éléments d_{ij} : des trajectoires de points ainsi que les vecteurs de flot optique. Nous notons les matrices correspondant à ces attributs SSM-pos and SSM-of respectivement.

Pour les trajectoires, les éléments d_{ij} sont exprimés comme la distance entre les positions des points qui sont suivis entre deux images. Considérons K points suivis avec succès entre les images I_i et I_j, le coefficient d_{ij} qui leur est associé dans la SSM, peut être calculé comme

$$d_{ij} = \sum_{k=1}^{K} \|\underline{x}_i^k - \underline{x}_j^k\|_2 \quad (2)$$

où $\underline{x}_i^k$ et $\underline{x}_j^k$ sont les positions du k^{eme} point dans les deux images. Nous obtenons les trajectoires en utilisant la méthode de suivi KLT [10, 11] : tout au long de la séquence, des points d'intérêts sont détectés et suivis pendant un certain temps. Ces trajectoires peuvent être courtes ou longues et correspondre à des points statiques ou mobiles. Dans le cas de mouvement de caméra, le mouvement apparent induit peut être estimé par un estimateur robuste [1, 6] et compensé de proche en proche de telle sorte que les coordonnées des points soient exprimées dans le repère de la première image de la séquence. En pratique, il peut y avoir au plus 500 points suivis au même instant c'est à dire pour une même image.
Pour calculer les vecteurs de flot optique, nous utilisons la méthode proposée par Lucas et Kanade [5]. Ces vecteurs sont utilisés pour calculer les éléments de la SSM comme suit :

$$d_{ij} = \sum_{k=1}^{n} \|\underline{d}_i^k - \underline{d}_j^k\|_2 \quad (3)$$

où $\underline{d}_i^k$ et $\underline{d}_j^k$ sont les vecteurs de mouvement du k^{eme} pixel dans chacune des deux images. En pratique le flot optique n'est pas calculé pour chaque pixel mais sur une grille, et la valeur n dépend par conséquent de la taille de l'image et de l'échantillonnage choisi pour la grille.

Descripteur local de SSM. Pour réaliser la synchronisation, il est nécessaire de caractériser la structure de la SSM. En effet, cette structure est stable sous des changements de vue. Nous choisissons la même représentation locale que dans [4] afin de décrire les matrices d'auto-similarités. Ce choix est motivé par le fait que l'incertitude des valeurs augmente avec la distance par rapport à la diagonale mais également par les délais et/ou déformations temporels qui peuvent influencer les structures de la matrice.

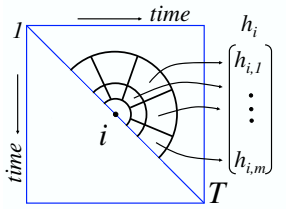


FIG. 1 – Les descripteurs locaux d'une SSM sont centrés sur chaque point de la diagonale $i = 1...T$ et reposent sur une structure log-polaire. Les histogrammes de directions du gradient sont calculés séparément pour chaque bloc et concaténés en un descripteur h_i.

Un descripteur local est calculé pour chaque élément diagonal. Nous construisons un histogramme de direction du gradient, normalisé à 8 classes, pour chacun des blocs de la structure log-polaire comme illustré à la Fig. 1. Le vecteur descripteur, h_i correspondant à l'image I_i, est obtenu en concaténant les histogrammes normalisés. Finalement, le descripteur temporel calculé pour une séquence d'images est une séquence de tels descripteurs $H = (h_1, ..., h_T)$. En pratique, le rayon de la structure log-polaire est d'environ 20 à 30 images.

2.2 Alignement des descripteurs

Cette section porte sur le problème d'alignement ou de synchronisation des descripteurs. En premier lieu, nous décrivons la méthode utilisée pour des séquences d'images sans *a priori* sur la fonction de déformation temporelle. Ensuite, nous proposons une approche plus simple quand cette fonction recherchée se réduit à un délai temporel constant.

Synchronisation sans *a priori* : DTW. Nous souhaitons aligner ou synchroniser les descripteurs extraits des matrices d'auto-similarités. Ce problème est similaire à celui qui consiste à aligner deux signaux temporels en reconnaissance de parole comme dans [7]. L'algorithme DTW est un outil classique pour résoudre ce type de problème. Dans notre cas, DTW est utilisé afin d'estimer la fonction de déformation w entre les axes temporels des deux séquences d'images. La correspondance entre les images d'indices i et j des deux séquences est exprimée par $j = w(i)$.
Soient deux séquences d'images I^1 et I^2 de longueurs N et M, nous calculons les SSMs et les descripteurs temporels correspondants, respectivement $H^1 = (h_1^1, ...h_i^1, ..., h_N^1)$ et $H^2 = (h_1^2, ...h_j^2, ..., h_M^2)$. Pour une mesure de dissimilarité S (plus la valeur de $S(h_i^1, h_j^2)$ est petite, plus la similarité entre h_i^1 et h_j^2 est grande), nous définissons la matrice de coût $\mathcal{C}$ comme

$$\mathcal{C} = [c_{ij}]_{i=1..N, j=1..M} = \left[S(h_i^1, h_j^2)\right]_{i=1..N, j=1..M}. \quad (4)$$

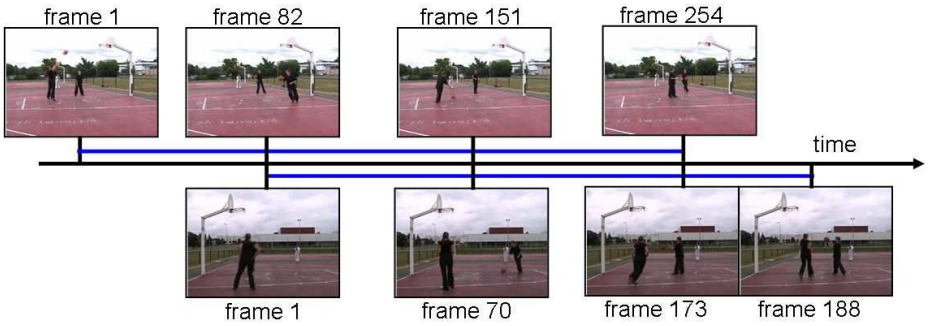
Chaque élément, c_{ij}, de cette matrice mesure le coût pour aligner la i^{eme} image de la première séquence avec la j^{eme} image de la seconde. Le meilleur alignement temporel est l'ensemble des paires $\{(i, j)\}$ qui contribuent au minimum global d'une mesure de dissimilarité cumulée. Le coût minimum cumulé, C_T, est

$$C_T = \min_w \sum_{i=1}^{N} S(h_i^1, h_{w(i)}^2). \quad (5)$$

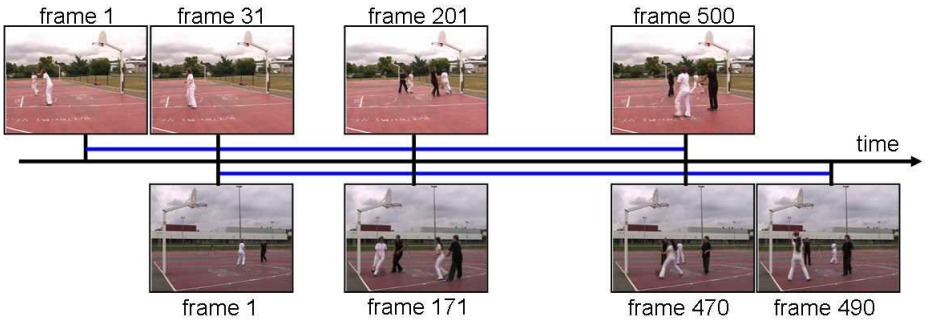
où la minimisation est réalisée dans l'ensemble des fonctions de déformations admissibles w. Pour résoudre (5) en utilisant la programmation dynamique, nous considérons 3 "déplacements" possibles (horizontal, vertical et diagonal) dans $\mathcal{C}$ pour les déformations admissibles. Nous pouvons calculer récursivement, pour chaque paire d'images (i, j), le coût minimum partiel cumulé par

$$\begin{aligned} C_A(h_i^1, h_j^2) &= c_{ij} + \min[C_A(h_{i-1}^1, h_j^2), \\ &\quad C_A(h_{i-1}^1, h_{j-1}^2), C_A(h_i^1, h_{j-1}^2)]. \quad (6) \end{aligned}$$

Les "déplacements" verticaux et horizontaux correspondent à l'association d'une image dans une séquence avec deux images consécutives dans l'autre séquence alors



(a)



(b)

FIG. 2 – Paires de séquences d'images de basket-ball. La paire située au-dessus présente des séquences montrant deux joueurs alors que la paire située au-dessous montre des séquences avec quatre joueurs.

qu'un "déplacement" diagonal consiste à associer deux paires d'images consécutives.
La solution finale C_T de (5) est par définition $C_T = C_A(h_N^1, h_M^2)$. Finalement, nous obtenons l'ensemble des paires $\{(i,j)\}$, représentant la fonction de déformation optimale w, en remontant depuis la paire (N,M), le chemin optimal dans la matrice de coût cumulé C_A.
Comme mentionné ci-dessus, une mesure de dissimilarité $S(\cdot,\cdot)$ est nécessaire afin d'évaluer de coût d'alignement. Nous choisissons simplement la distance euclidienne entre les descripteurs H^1 and H^2.

Synchronisation pour un délai temporel fixe. L'approche proposée dans ce paragraphe peut être employée lorsque la fonction de déformation est restreinte à un délai temporel fixe inconnu. Comme auparavant, nous calculons les SSMs et les descripteurs pour les deux séquences d'images et la matrice de coût correspondant, $\mathcal{C}$, définie par (4).
Pour chaque valeur entière possible de délai temporel, dt, nous calculons le coût moyen

$$c(dt) = \frac{\sum_{i=\max(1,1-dt)}^{\min(N-dt,M)} S(h_i^1, h_{i+dt}^2)}{\min(N-dt, M) - \max(1, 1-dt) + 1}. \quad (7)$$

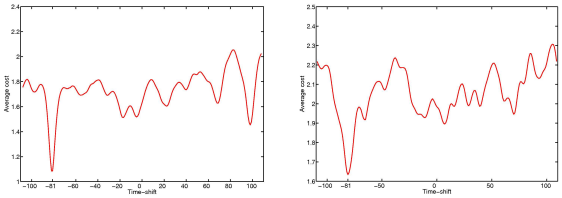
Nous pouvons alors tracer ce coût moyen c en fonction du délai temporel dt. Cette représentation est similaire à celle proposée par [16] où une mesure d'énergie extraite à partir de contraintes de rang est évaluée en fonction du délai temporel. En pratique, nous traçons la fonction c pour dt appartenant à l'intervalle $[-\frac{M+N}{4}, \frac{M+N}{4}]$.

3 Résultats de synchronisation pour des scènes dynamiques

Dans cette section, nous présentons différents résultats de synchronisation de scènes dynamiques. Les premiers résultats, proposés au paragraphe 3.1, correspondent à des séquences d'images acquises à l'aide de caméras statiques alors que ceux du paragraphe 3.2 concernent des séquences captées par des caméras mobiles.

3.1 Caméras statiques

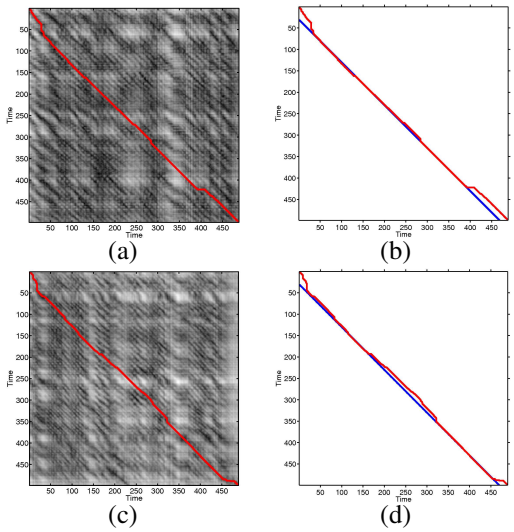
En premier lieu, nous validons la méthode de synchronisation de vidéos en considérant des séquences d'images de basket-ball enregistrées avec des caméras statiques. Dans la première paire de séquences, illustrée en Fig. 2(a), nous pouvons voir deux joueurs, alors que la seconde paire, illustrée en Fig. 2(b), montre quatre joueurs. Dans les deux paires, les points de vue des caméras sont presque opposés ce qui signifie qu'aucun point du fond n'est simultanément visible dans les deux vues. Nous calculons les SSM-of et SSM-pos ainsi que les descripteurs correspondants pour chacune des séquences d'images. Nous alignons ensuite les descripteurs relatifs à chaque type de matrice.
En supposant que la fonction de désynchronisation est un délai temporel constant, nous calculons, pour la première paire, le coût moyen comme fonction du délai. Nous traçons cette fonction pour les deux types de SSM comme illustré en Fig. 3. Nous pouvons observer que le minimum est atteint pour un délai de -81 images dans les deux cas ce qui correspond au délai réel.



(a) (b)

FIG. 3 – Résultats de synchronisation pour la scène de basket-ball avec deux joueurs avec l'hypothèse d'un délai temporel constant (a) Coût moyen du chemin en fonction du délai pour SSM-of (b) pour SSM-pos. Dans les deux cas, le chemin de coût moyen minimum est obtenu pour le délai réel -81.

En supposant que nous ne disposons d'aucun *a priori* concernant la fonction de désynchronisation, nous appliquons la méthode DTW pour les deux types de SSM dans le cas de la scène avec quatre joueurs. Les fonctions de mise en correspondances temporelles obtenues (courbes rouges en Fig. 4(b,d)) correspondent bien à la vérité terrain (courbes bleues) pour les deux types de SSM.

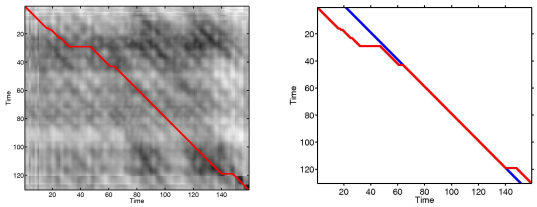


(a) (b) (c) (d)

FIG. 4 – Résultats de synchronisation de la scène de basket-ball avec quatre joueurs (a,b)Transformation temporelle (rouge) obtenue avec SSM-of, tracée sur la matrice de coût et la vérité terrain (bleue) (c,d) Résultats obtenus avec SSM-pos. Dans les deux cas, les estimations coïncident presque entièrement avec la vérité terrain (courbe bleue).

3.2 Caméras mobiles

Les deux méthodes peuvent être utilisées dans le cas de séquences d'images enregistrées par des caméras mobiles. Pour cela, considérons une paire de vidéos de football (Fig. 5(a)). Des trajectoires obtenues par la méthode KLT de suivi de points sont utilisées pour calculer les SSM-pos. Les résultats obtenus sous hypothèse de délai temporel fixe sont présentés en Fig. 5(b). Ceux obtenus en l'absence de cette hypothèse sont présentés en Fig. 6. A nouveau les résultats sont concluants.



(a) (b)

FIG. 6 – Résultats de synchronisation pour les séquences de football acquises en caméras mobiles et utilisant la méthode DTW (a) Matrice de coût avec l'estimation de la fonction de déformation (b) Cette estimation (courbe rouge) recouvre presque la vérité terrain (courbe bleue).

4 Résultats de Synchronisation pour des Actions Humaines

Dans cette Section, nous proposons des résultats de synchronisation d'actions humaines. Les premiers résultats ont été obtenus sur une base de données publique dédiée à la reconnaissance d'actions [15]. Ensuite le paragraphe 4.2 est consacré à des actions naturelles extraites de différentes vidéos.

4.1 Base d'actions IXMAS

En premier lieu, nous avons tenté d'aligner temporellement des séquences d'actions extraites de la base "IXMAS"[1]. Cette base est composée de 5 vues différentes de 10 acteurs réalisant chacune des 11 classes d'actions différentes trois fois. Les positions ainsi que les orientations sont choisies arbitrairement par les acteurs.

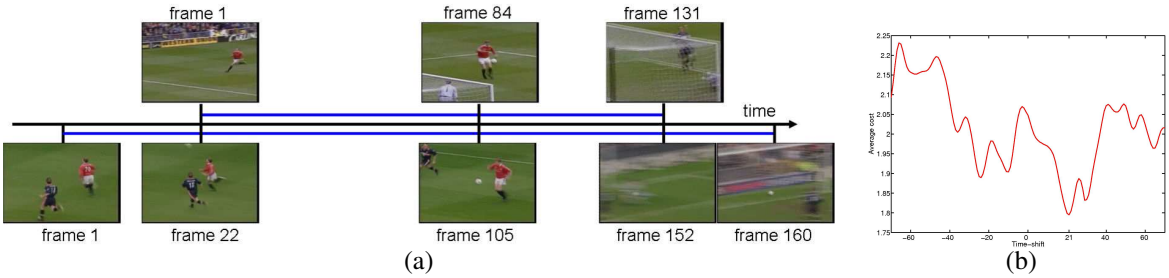
Pour ces actions, nous alignons les séquences d'images à partir des descripteurs de SSM-of. Pour chaque action, le flot optique est calculé sur des boîtes englobantes autour des acteurs. Ces boîtes sont extraites à partir des silhouettes disponibles pour chaque image. Nous présentons un exemple de synchronisation pour l'action "checkwatch"réalisées par deux personnes différentes et considérant des vues très différentes : une vue de côté et une vue de dessus. Le résultat est illustré en Fig. 7. Les correspondances entre images de Fig. 7(b) montrent que l'alignement temporel est correct malgré les différences de durées entre ces deux réalisations.

4.2 Actions naturelles

Pour finir, des tests ont également été menés sur des actions extraites de séquences d'images naturelles.

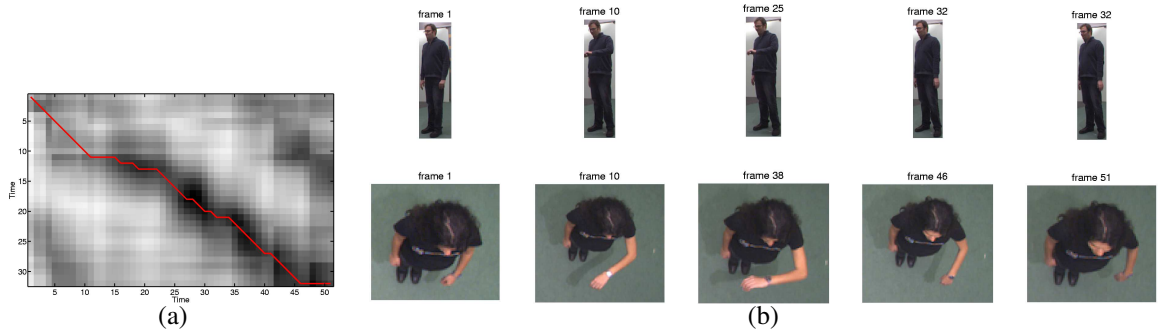
Tir de basket-ball. Dans les scènes de basket-ball précédemment évoquées, nous disposons de plusieurs tirs réalisés par différents joueurs et vus sous différents points de vue. Nous alignons ces actions de tirs à partir des descripteurs de SSMs-of. Les vecteurs de flot optique sont calculés sur des boîtes englobantes centrées sur le joueur. Ces boîtes englobantes ont été annotées manuellement. Un exemple

[1] https ://charibdis.inrialpes.fr/html/sequences.php



(a) (b)

FIG. 5 – Séquences de football acquises par des caméras mobiles (a) Axes temporels des deux séquences représentés par quelques images (b) Coût moyen du chemin en fonction du délai temporel. Le minimum est atteint pour la valeur réelle 21.



(a) (b)

FIG. 7 – Synchronisation de deux actions “check-watch”. (a) Matrice de coût et chemin estimé. (b) Quelques correspondances sur ce chemin.

d’une telle synchronisation est présenté en Fig. 8. L’alignement temporel retrouvé est encore correct malgré les importantes différences dans la durée et la réalisation des deux tirs.

Actions “Boire” et “Fumer” dans un film. Dans le film *Coffee and Cigarettes*, les actions “boire” et “fumer” sont réalisées de nombreuses fois par différents acteurs et sous différents points de vue. Pour ces deux types d’actions, nous calculons les SSM-of. Les vecteurs de flot optique sont calculés sur des boîtes englobantes extraites à partir d’annotations. Nous obtenons de bons résultats, comme l’illustrent les Fig. 9 et Fig. 10, pour ces deux types d’actions. Nous pouvons observer, en Fig 9(a) et Fig 10(a), que ces séquences ont des longueurs différentes ce qui signifie que les actions sont réalisées à des vitesses différentes. En Fig 9(b) et Fig 10(b), les correspondances temporelles obtenues entre les deux séquences sont illustrées par plusieurs couples $(I_i^1, I_{w(i)}^2)$.

5 Conclusion

Nous avons présenté un cadre commun pour la synchronisation d’actions humaines et de scènes dynamiques en présence de changements importants de point de vue. Cette méthode repose sur les similarités et dissimilarités temporelles dans les séquences d’images. Comme les structures des matrices d’auto-similarité sont à la fois stables pour des points de vue différents et caractéristiques pour des classes de motifs dynamiques, elles constituent un descripteur temporel approprié pour l’alignement des séquences d’images dans ces deux contextes. Un des principaux avantages de cette méthode est qu’elle ne repose pas sur des hypothèses restrictives telles que l’existence de correspondances de points entre les vues ou encore la présence, dans le fond, d’informations visuelles suffisantes. Grâce à l’utilisation de DTW, nous pouvons réaliser les deux tâches même lorsque le désalignement temporel n’est pas un simple délai, mais une transformation monotone arbitraire. Nous avons testé les performances de notre approche sur des séquences d’images réelles acquises par des caméras statiques et *mobiles*.

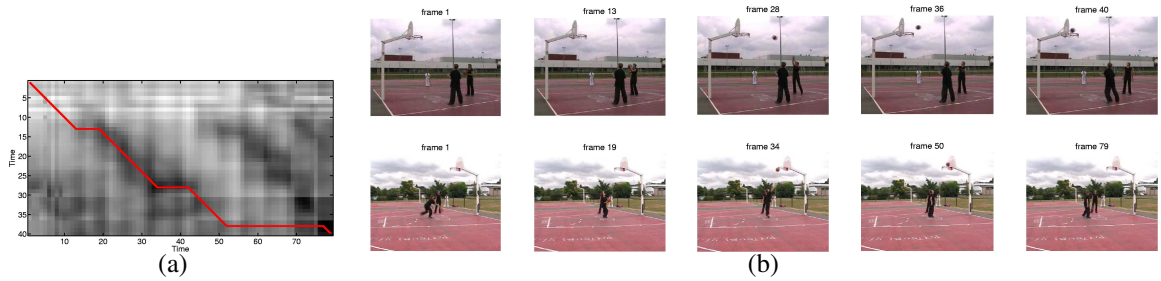
Dans ces travaux, nous supposons que les deux séquences d’images correspondent à un évènement dynamique identique ou similaire. Cependant, la méthode pourrait être exploitée dans des travaux futurs à des fins de regroupement d’actions, de détection d’actions, de détection de vidéos ou encore de mise en correspondance de vidéos.

Références

[1] M. J. Black and P.Chau Anandan. The robust estimation of multiple motions : Parametric and piecewise-smooth flow fields. *Computer Vision and Image Understanding*, 63(1) :75–104, January 1996.

[2] R.L. Carceroni, F.L.C. Padua, G.A.M.R. Santos, and K.N. Kutulakos. Linear sequence-to-sequence alignment. In *Proc. Conf. Comp. Vision Pattern Rec.*, pages I : 746–753, 2004.

[3] Y. Caspi and M. Irani. Spatio-temporal alignment of sequences. *IEEE Trans. on Pattern Anal. and Machine Intell.*, 24(11) :1409–1424, November 2002.



(a) (b)

FIG. 8 – Synchronisation de deux tirs de basket-ball. (a) Matrice de coût et chemin estimé. (b) Quelques correspondances sur ce chemin.

(a) (b)

FIG. 9 – Synchronisation de deux actions “boire”.(a) Matrice de coût et chemin estimé. (b) Quelques correspondances sur ce chemin.

(a) (b)

FIG. 10 – Synchronisation de deux actions “fumer”.(a) Matrice de coût et chemin estimé. (b) Quelques correspondances sur ce chemin.

[4] I.N. Junejo, E. Dexter, I. Laptev, and P. Pérez. Cross-view action recognition from temporal self-similarities. In *Proc. Eur. Conf. Comp. Vision*, pages 293–306, 2008.

[5] B.D. Lucas and T. Kanade. An iterative image registration technique with an application to stereo vision. In *Image Understanding Workshop*, pages 121–130, 1981.

[6] J.-M. Odobez and P. Bouthemy. Robust multiresolution estimation of parametric motion models. *Journal of Visual Communication and Image Representation*, 6(4) :348–365, December 1995.

[7] L. Rabiner, A. Rosenberg, and S. Levinson. Considerations in dynamic time warping algorithms for discrete word recognition. *IEEE Trans. on Acoustics, Speech and Signal Processing*, 26(6) :575– 582, 1978.

[8] A. Rao, C.and Gritai, M. Shah, and T. F. Syeda Mahmood. View-invariant alignment and matching of video sequences. In *Proc. Int. Conf. on Image Processing*, pages 939–945, 2003.

[9] G.P. Stein. Tracking from multiple view points : Self-calibration of space and time. In *Proc. Conf. Comp. Vision Pattern Rec.*, volume 1, pages 521–527, 1999.

[10] C. Tomasi and T. Kanade. Detection and tracking of point features. Technical report, Carnegie Mellon University Technical Report CMU-CS-91-132, 1991.

[11] C. Tomasi and J. Shi. Good features to track. In *Proc. Conf. Comp. Vision Pattern Rec*, pages 593–600, 1994.

[12] T. Tuytelaars and L.J. Van Gool. Synchronizing video sequences. In *Proc. Conf. Comp. Vision Pattern Rec.*, volume 1, pages 762–768, 2004.

[13] Y. Ukrainitz and M. Irani. Aligning sequences and actions by minimizing space-time correlations. In *Proc. Europ. Conf. on Computer Vision*, 2006.

[14] M. Ushizaki, T. Okatani, and K. Deguchi. Video synchronization based on co-occurrence of appearance changes in video sequences. In *Int. Conf. on Pattern Recognition*, pages III : 71–74, 2006.

[15] D. Weinland, E. Boyer, and R. Ronfard. Action recognition from arbitrary views using 3d exemplars. In *Proc. Int.Conf. on Computer Vision*, pages 1–7, 2007.

[16] L. Wolf and A. Zomet. Wide baseline matching between unsynchronized video sequences. *Int. J. of Computer Vision*, 68(1) :43–52, June 2006.